

Contents

<i>Preface</i>	<i>page</i> ix
<i>Nomenclature</i>	xii
Part I Formulation	1
1 Bayesian Learning	3
1.1 Framework	3
1.2 Computation	10
2 Variational Bayesian Learning	39
2.1 Framework	39
2.2 Other Approximation Methods	51
Part II Algorithm	61
3 VB Algorithm for Multilinear Models	63
3.1 Matrix Factorization	63
3.2 Matrix Factorization with Missing Entries	74
3.3 Tensor Factorization	80
3.4 Low-Rank Subspace Clustering	87
3.5 Sparse Additive Matrix Factorization	93
4 VB Algorithm for Latent Variable Models	103
4.1 Finite Mixture Models	103
4.2 Other Latent Variable Models	115
5 VB Algorithm under No Conjugacy	132
5.1 Logistic Regression	132
5.2 Sparsity-Inducing Prior	135
5.3 Unified Approach by Local VB Bounds	137

Part III	Nonasymptotic Theory	147
6	Global VB Solution of Fully Observed Matrix Factorization	149
6.1	Problem Description	150
6.2	Conditions for VB Solutions	152
6.3	Irrelevant Degrees of Freedom	153
6.4	Proof of Theorem 6.4	157
6.5	Problem Decomposition	160
6.6	Analytic Form of Global VB Solution	162
6.7	Proofs of Theorem 6.7 and Corollary 6.8	163
6.8	Analytic Form of Global Empirical VB Solution	171
6.9	Proof of Theorem 6.13	173
6.10	Summary of Intermediate Results	180
7	Model-Induced Regularization and Sparsity Inducing Mechanism	184
7.1	VB Solutions for Special Cases	184
7.2	Posteriors and Estimators in a One-Dimensional Case	187
7.3	Model-Induced Regularization	195
7.4	Phase Transition in VB Learning	202
7.5	Factorization as ARD Model	204
8	Performance Analysis of VB Matrix Factorization	205
8.1	Objective Function for Noise Variance Estimation	205
8.2	Bounds of Noise Variance Estimator	207
8.3	Proofs of Theorem 8.2 and Corollary 8.3	209
8.4	Performance Analysis	214
8.5	Numerical Verification	228
8.6	Comparison with Laplace Approximation	230
8.7	Optimality in Large-Scale Limit	232
9	Global Solver for Matrix Factorization	236
9.1	Global VB Solver for Fully Observed MF	236
9.2	Global EVB Solver for Fully Observed MF	238
9.3	Empirical Comparison with the Standard VB Algorithm	242
9.4	Extension to Nonconjugate MF with Missing Entries	247
10	Global Solver for Low-Rank Subspace Clustering	255
10.1	Problem Description	255
10.2	Conditions for VB Solutions	258
10.3	Irrelevant Degrees of Freedom	259
10.4	Proof of Theorem 10.2	259

Contents	vii
10.5 Exact Global VB Solver (EGVBS)	264
10.6 Approximate Global VB Solver (AGVBS)	267
10.7 Proof of Theorem 10.7	270
10.8 Empirical Evaluation	274
11 Efficient Solver for Sparse Additive Matrix Factorization	279
11.1 Problem Description	279
11.2 Efficient Algorithm for SAMF	282
11.3 Experimental Results	284
12 MAP and Partially Bayesian Learning	294
12.1 Theoretical Analysis in Fully Observed MF	295
12.2 More General Cases	329
12.3 Experimental Results	332
Part IV Asymptotic Theory	339
13 Asymptotic Learning Theory	341
13.1 Statistical Learning Machines	341
13.2 Basic Tools for Asymptotic Analysis	344
13.3 Target Quantities	346
13.4 Asymptotic Learning Theory for Regular Models	351
13.5 Asymptotic Learning Theory for Singular Models	366
13.6 Asymptotic Learning Theory for VB Learning	382
14 Asymptotic VB Theory of Reduced Rank Regression	385
14.1 Reduced Rank Regression	385
14.2 Generalization Properties	396
14.3 Insights into VB Learning	426
15 Asymptotic VB Theory of Mixture Models	429
15.1 Basic Lemmas	429
15.2 Mixture of Gaussians	434
15.3 Mixture of Exponential Family Distributions	443
15.4 Mixture of Bernoulli with Deterministic Components	451
16 Asymptotic VB Theory of Other Latent Variable Models	455
16.1 Bayesian Networks	455
16.2 Hidden Markov Models	461
16.3 Probabilistic Context-Free Grammar	466
16.4 Latent Dirichlet Allocation	470

17 Unified Theory for Latent Variable Models	500
17.1 Local Latent Variable Model	500
17.2 Asymptotic Upper-Bound for VB Free Energy	504
17.3 Example: Average VB Free Energy of Gaussian Mixture Model	507
17.4 Free Energy and Generalization Error	511
17.5 Relation to Other Analyses	513
<i>Appendix A James–Stein Estimator</i>	516
<i>Appendix B Metric in Parameter Space</i>	520
<i>Appendix C Detailed Description of Overlap Method</i>	525
<i>Appendix D Optimality of Bayesian Learning</i>	527
<i>Bibliography</i>	529
<i>Subject Index</i>	540